Master's Thesis

Neuron segmentation using incomplete and noisy labels via adaptive learning with structure priors

ChanMin Park

Department of Computer Science and Engineering

College of Information-Bio Convergence Engineering

2021

Neuron segmentation using incomplete and noisy labels via adaptive learning with structure priors

ChanMin Park

Department of Computer Science and Engineering

College of Information-Bio Convergence Engineering

Neuron segmentation using incomplete and noisy labels via adaptive learning with structure priors

A thesis/dissertation submitted to
Ulsan National Institute of Science and Technology
in partial fulfillment of the
requirements for the degree of
Master of Science

Chan-Min Park

12/15/2020

Approved by

Advisor
Se Young Chun

Neuron segmentation using incomplete and noisy labels via adaptive learning with structure priors

Chan-Min Park

This certifies that the thesis/dissertation of Chan-Min Park is approved.

12/15/2020 of submission

Signature

Advisor: Se Young Chun

Signature

Won-Ki Jeong: Thesis Committee Member

Signature

Jae-Young Sim: Thesis Committee Member

Abstract

Recent advances in machine learning have demonstrated significant success in biomedical image segmentation. Most existing high-quality segmentation algorithms rely on supervised learning with full training labels. However, segmentation is more susceptible to label quality; notably, generating accurate labels in biomedical data is a labor- and time-intensive task. Especially, structure neuronal images are hard to obtain full annotation because of the entangled shape of each structure. In this thesis, a neuron structure semantic segmentation algorithm is proposed on a noise label. I assume that the label has noise and propose two new novel loss functions. Adaptive loss is applied to noise pixels in different labels with prediction in partially annotated labels. These fluorescence images may have confidence that can leverage prior knowledge when each pixel has intensity. Reconstruction loss is suggested that can be regularized of neuronal cell structures to reduce false segmentation near noisy labels. Additionally, This study is aimed to verify that our method preserves the connectivity of linear structure through a novel evaluation matrix.

Contents

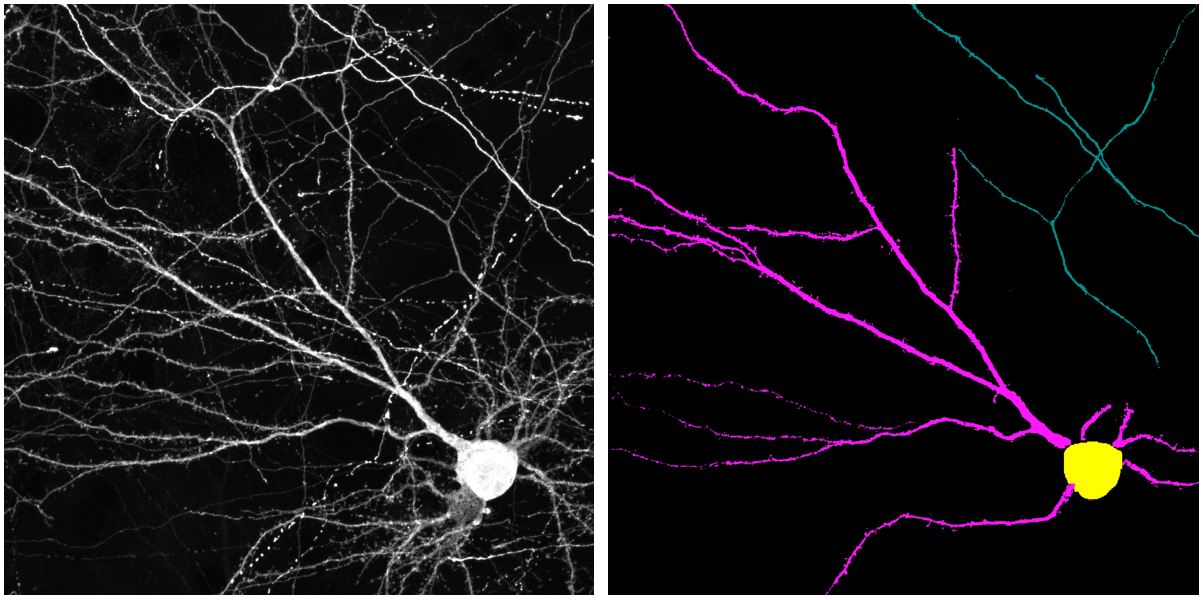
List of Figures

I Introduction

1.1 Background

Neurons show highly polarized structures that typically consist of multiple dendrites and a single axon from a cell body. The morphology of a neuron is critical for various neuronal properties, including synaptic integration and neuronal excitability [1]. Moreover, abnormalities of neuronal shapes have been observed in multiple neurodegenerative diseases, such as Alzheimer's disease and Parkinson's disease [2]. Recent advances in imaging techniques led to the emergence of large scale image datasets of complex neuronal processes which require future analysis of automated and precise analysis tools. Currently, commercially available software and open-source tools are broadly applied for segmenting and annotating neurons [3, 4]; however, there is still a room for improvement in labeling accuracy and automation of the process.

Many semantic segmentation methods rely on supervised learning with clean labels [5–7]. However, this data (fluorescence microscopy images of neurons) are different from those in conventional semantic segmentation problems; neurons are narrow and thin, and they are distributed sparsely (i.e., a large portion of the image contains empty background). Moreover, multiple neurons are intertwined and cross each other. Making pixel-level accurate ground-truth labels from neuron images is laborious and difficult. The input to our method is partially labeled data (see Figure 1), that is some axons and dendrites are marked, while the others are left unmarked; in other words, weak labels. We also consider these noisy labels because unlabeled pixels work negatively during training when the loss function is the conventional Mean Squared Error (MSE). Recently, many noise-robust loss functions have been proposed, such as Mean Absolute Error (MAE) [8], Symmetric cross entropy Learning(SL), Reverse Cross-Entropy (RCE) [9], Normalized Cross-Entropy (NCE), Active and Passive Loss (APL) [10], and Noise-Robust Dice (NR-Dice) [11]. However, none of them has shown satisfactory performance regrading our target problem.



(a) Input (b) Partial label

Figure 1: Our neuron data are composed of (a) input fluorescence image showing complex structures, and (b) partially given labels (magenta: dendrites, teal green: axons).

1.2 Motivation

Some studies approach the unlabeled pixel as weakly supervised, but I try to approach the problem by considering the unlabeled part to constitute noise. In the case of general cross-entropy, negative log-likelihood is used for the noise part at 0. Therefore, the unlabeled pixel does not contribute to learning, thus, there is not enough data for training. In addition, cross-entropy is more susceptible to noise in multi-class task because there are potentially more varied values. However, in the case of MSE and MAE, these are robust against noise because the noise parts also contribute to learning [12]. The MAE was theoretically proved that work well against noise such as symmetric, asymmetric, but it takes a lot of time to stabilize. Although MSE, which is bounded in MAE, is not strong against all kinds of noise, learning is fast due to square values. [8] Therefore, I will use MSE to contributing to the noise pixel.

In addition, networks are needed to selectively back-propagate according to the presence or absence of noise. Multi-task learning can inspire that this approach can apply different loss when the possibility of applying other losses [13]. Nevertheless, it is difficult for the network to classify the unlabeled parts during training because the predicted values (that is, the axon and dendrite parts) are similar. The context prediction [14] method inspire to me by using self-supervised learning. The foreground image(right images(a) in Figure 1 in the fluorescence images can be influenced by the technique of classifying each pixel using the contextual relationship between prediction results. The position between pixels can be classified through the regularization of predicted values. This approach relies on the assumption that the positions between pixels can be classified through normalization of each other. Therefore, since one pixel in the foreground

should have one class, we adopted with self-regularization so that it is assigned to the final class by measuring the relationship between the predicted channels.

1.3 Problem definition

When label is incomplete, or noise pixels, segmentation performance will be degraded. Many segmentation methods ignored noise pixel when training or remove noise data from preprocessing. However, incomplete pixel may have important characteristics. Also, in this data, there is a lot of information about unlabeled pixel. I thought how incomplete pixels is classified, if only using the incomplete or noise label. As shown in Figure 2, partial label will be utilized during the training with proposed my loss and full label will be evaluated in testing.

1.4 Contributions

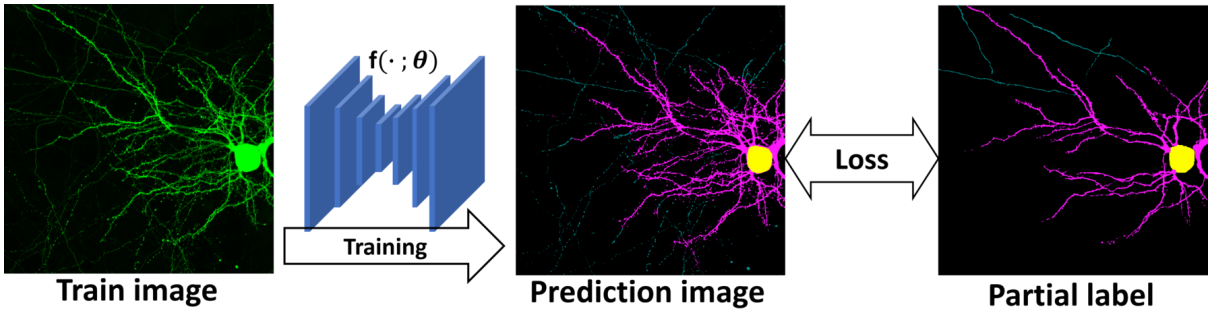
In this paper, two novel loss functions are proposed that are specifically designed to handle incomplete and noisy labels for neuron segmentation. Our method is also shown to how much improve the performance of neuron segmentation using the center line Dice(ClDice) metric [15], which measures the connectivity of elongated structures.

- Adaptive mean squared error (ADMSE) loss is developed, which employs spatially varying weights to prevent learning from noisy labels.
- Moreover, a structure prior (STPR) loss is proposed to promote the correct assignment of labels when multiple classes compete with each other.
- The performance of proposed method demonstrate by comparing it with other noise-robust losses using various segmentation accuracy metrics.

1.5 Goal

As shown in Figure 2, the two process will be examined in this study. In the first process, only partial labeling will be used during training, and the optimal parameter-based best model is evaluated in the test process. This process will ensure that only a partial label can provide the full label with the correct class for each pixel. It is expected that the structural characteristics of the neuron can be obtained, as well as improvement of the linear structure.

(a) Training process



(b) Test process

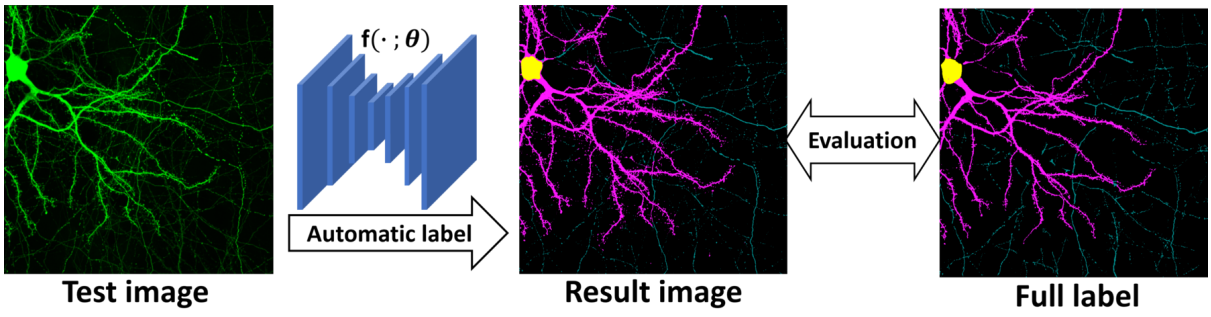


Figure 2: This shows our overview process for each (a)training and (b)test processes. The first (a) process only used partial labels, and then the best model evaluates scores in the (b) process.

II Related work

2.1 Research of neuron structure

In general, the structure of neurons in microscopic images plays a very important role in understanding function of the brain. In general, studies of neuron have focused on for restoration to get a neuronal structure. Ongoing studies mainly focus on neuron 3D reconstruction or analyzing the structure of neurons using tracing technology. In order to apply tracing and reconstruction, accurate segmentation of single neurons is required. The previous method used has a problem in its mathematical modeling [16]. The previous method is difficult to work with when multiple neurons are attached, and it is hard to segment each image precisely. Recently, deep learning has been used to study the segmentation of neuron structures. Zhi at al, [17] presented a deep learning-based open tracing tool box. They suggested a user-interactive method of pruning neurons or improving the quality of their morphology with semi-automatically segmentation. Therefore, many studies are currently actively approaching segmentation based on deep learning; in addition, the DIADEM challenge and the BigNeuron project [18] have provided a large amount of data and tracing labels.

Recently, research has shown that the a 3D binary mask obtained by applying 3D-Unet to a single neuron structure is better than the results of tracing on the original image [19]. Furthermore, in this method, a single neuron is segmented from the overlapping neuron structure based on a weakly label, producing improved results without intersecting points in tracing [20]. Nevertheless, it is still difficult to distinguish between axon and dendrite as the results of tracing. Because the neuron is composed of single structure, that is the entire structure of the neuron. In this study, a new approach is presented with semantic segmentation of the axon, dendrite, and cell body that has not been previously developed. Through the proposed my method, the characteristics and length of neurons in each axon and dendrite structure can be obtained. In addition, many organic substances in the neuron structure can be analyzed for each structure to find the differences and present a new research approach.

2.2 Semantic segmentation

The rapid development of deep learning in image processing has been applied in various fields, such as segmentation, detection, and restoration. Of these, image segmentation is used in various fields such as, disease detection and autonomous driving [21]. Semantic segmentation has been used to determine accurate class by classifying each pixel unit. The algorithm can implemented through various networks, such as Convolution Neural Network(CNN) and Recurrent Neural Network(RNN) [22], and can utilize with diverse dataset; for example, Pascal VOC 2012 [23], COCO datasets [24], and Cityscape [25] have been used to achieve consistently high performance. The network is gradually improving performance through these various structures. The Fully Convolution Network(FCN) [26] consists of an end-to-end-based, shape such as VGG16 and GoogleNet. In the encoder part of the preceding network, there are Resnets [27] and DenseNets [28] that make features more deep, and ResNeXt [29], which improves performance by making complex. Furthermore, a receptive field in the convolution activation field was studied by researcher. Then the DeepLab V3+ [30] was developed based on Atrous convolution. In addition, it provides a decoder area to show the performance as a deconvolution base. U-net was also used as a representative in medical imaging, and gradient vanishing was avoided by using skip connections. Likewise, it is used as a Generative Adversarial Network(GAN) to approach the problem.

The prediction map is extracted using a generator for general data, and the image is divided by classifying it using the discriminator. In addition, the decoder part provides deconvolution to demonstrate the improved performance [31,32]. The aforementioned methods have been studied to, but in recent years, research has been conducted with weakly and semi-supervised learning rather than through supervised learning for semantic segmentation. There is a Class Activation Mapping(CAM) that uses an activation map through a label for an image classification [33, 34]. Additionally, iterative segmentation has been developed by applying seed growing technology [35] in weakly supervised learning. Other methods use little information, such as scribbles [36] or points. In recent years, image segmentation has been studied significantly in medical imaging. Since it is difficult to determine accurate labels due to the nature of the type of image, the most recent approaches are weakly and semi-supervised to solve this problem. In this thesis, I will approach the image problem using the segmentation method.

2.3 Noise label learning

Presently, many deep learning approaches depended on the noise label. If damaged labels were included in the dataset, network would have a problem of overfitting, which would lead to a decrease in overall performance. Some studies have approached this problem by suggesting various methods in noise perspective. There are many approaches, such as pre-processing data, making noise-robust architecture responded to noise, or regularization. [37] As in data pre-possessing through the noise-robust method, there is a simple method to remove data when has the noise that influences training through using bagging and boosting. Recently, a method of noise pruning has been developed that does not require parameters called confident learning. It utilizes joint distribution with unknown noise and prediction, and removes noise according to rank [38]. From the perspective of network architecture, training is sometimes performed by adding an adaptive layer that prevents noise when the predicted value is lower than an arbitrary value. In order for the network to be robust against noise labels, data augmentation, weight day, or dropout can be used to create a tolerance to noise, which makes it more robust to classify. Recently, there a method has been introduced to learn about noise by performing label smoothing and giving a more label with noise; this is termed the Mixup method [39].

2.4 Noise-robust loss

Many methods have been described above. In this study, I will approach the noise loss more specifically. In general, the goal of noise-based learning is to create a function $f(\mathbf{X};\theta)$ that matches the label($\mathbf{Y}$) with the prediction value($\hat{\mathbf{Y}}$). At this time, the risk of the model f is defined as the expected value of the loss function(L). The process of finding the optimal f that minimizes to zero the expected value of the loss is termed Empirical Risk Minimization(ERM). Let define ERM in the dataset(D), followed by

$$\mathcal{R}(f) = \mathbb{E}_D \left[L(f(\mathbf{X};\theta), \mathbf{Y} \right] \tag{1}$$

Previously, in binary classification problems involving noise, 0-1 loss was theoretically proven to be a robust method. This simply loss function that gives a penalty of 0 if the predicted value and label are accurate. However, since 0-1 loss is not applied to noise even in multi-class classification problems, Ghosh et al. [8] theoretically demonstrated the validity of suitable loss MAE. When input $\mathbf{X}$ has the k class, Let denotes that the ground-truth distribution is $q(k|\mathbf{X})$ and each label probability is $p(k|\mathbf{X})$. MSE satisfied the ERM(eq. 1) in symmetric and asymmetric noise and is defined as :

$$MAE = \sum_{k=1}^{K} |p(k|\mathbf{X}) - q(k|\mathbf{X})| \tag{2}$$

In addition, it was shown that works well for noise, as much as $\frac{k}{k-1}$ noise ratio. However, in the case of MAE, as the ratio of noise increases, it takes a lot of time to saturate the model. To over-come this problem, cross entropy(CE) recently is developed by a symmetric cross entropy(SCE)

function that is robust to noise in multi-class classification. [9] This function is motivated by KL-divergence, and these function defined as:

$$CE = -\sum_{k=1}^{K} q(k|\mathbf{X}) log(p(k|\mathbf{X})) \tag{3}$$

$$RCE = -\sum_{k=1}^{K} p(k|\mathbf{X}) log(q(k|\mathbf{X})). \tag{4}$$

In a similar method, there is a generalized cross-entropy [12] that forms a Box Cox shape by combining MAE with CE. If the predicted value is smaller than a certain value, the network is applied to the noise robust loss. This loss only operates in the specific area with noise; thus, Xingjun at al. [10] found that many classification systems have the problem of under-fitting. Therefore, they suggest normalized active and passive loss that combines normalized cross-entropy and normalized noise-robust loss, theoretically implementing robustness against noise. Recently, NR-DICE [11] showed good performance by using the loss combined with Dice loss and MAE in medical images. However, there are still many problems because noise is irregularly generated in medical images. Therefore, in this study, I propose a method that is robust to noise in neuronal images derived from fluorescence images.

III Method

3.1 Overview

Let $\mathcal{X}$ and $\mathcal{Y}$ be the distributions of observed images and neuron labels, respectively. neuron segmentation is set goal to predict a response map $\mathbf{Y} \sim \mathcal{Y}$ for every class representing neuronal structures from a sample image $\mathbf{X} \sim \mathcal{X}$ by feeding a deep neural network $f(\cdot;\theta)$ parameterized by θ using a paired random variable $(\mathbf{X}, \mathbf{Y})$ from the distribution $\mathcal{D}$. In our problem, four classes are used , such as background, cell body, dendrite, and axon, which are denoted as labels 0, 1, 2, and 3 respectively, as shown in Figure 3. Since the labels are partially given, all unlabeled pixels cannot be treated as background. Instead, an intensity-based thresholding is applied to the input image to generate the background image $\mathbf{Y}^{(0)}$. K is defined as the set of label indices (in our case, $K = \{0, 1, 2, 3\}$). Note that the cell body and background labels are considered as clean labels (i.e., not partial labels). Therefore, the set of clean labels are defined $S_1 = \{0, 1\}$ and that of noisy labels $S_2 = \{2, 3\}$, and thus, $K = S_1 \cup S_2$.

3.2 ADMSE loss

The MSE loss is widely used to minimize the distance between the prediction $f(\mathbf{X};\theta)$ and the ground-truth label $\mathbf{Y}$ as follows:

$$\mathcal{L}_{mse} = \min_{\theta} \mathbb{E}_{(\mathbf{X},\mathbf{Y})\sim\mathcal{D}} ||f(\mathbf{X};\theta) - \mathbf{Y}||^2 \tag{5}$$

Even though the MSE loss effectively penalizes large errors in L^2 sense, it is not resilient to noise in training data as in our problem. To address this issue, an adaptive weight are proposed based on the prediction of the network to control the pixel-level backpropagation. Let $\mathbf{Y}^{(k)}$ be the label map for the k-th class and $Y_i^{(k)}$ be the i-th pixel value (i.e., per-pixel label) in $\mathbf{Y}^{(k)}$. Then, L_{admse} be defined using a per-pixel adaptive weight function $w(\cdot,\cdot)$ as follows:

$$\mathcal{L}_{admse} = \mathbb{E}_{(\mathbf{X},\mathbf{Y})\sim\mathcal{D}}[\sum_{k}^{K}\sum_{i}^{I} w(\hat{Y}_i^{(k)}, Y_i^{(k)})(\hat{Y}_i^{(k)} - Y_i^{(k)})^2] \tag{6}$$

$$w(\hat{Y}_i^{(k)}, Y_i^{(k)}) = \begin{cases} e^{-(\frac{\hat{Y}_i^{(k)}}{\alpha})^2} + \beta & \text{if } Y_i^{(k)} = 0, k \in S_2 \\ 1 & \text{otherwise} \end{cases} \tag{7}$$

The rationale behind the proposed loss is as follows; if an unlabeled pixel ($Y_i^{(k)} = 0$) in the noisy label map ($k \in S_2$) has a higher prediction value $\hat{Y}_i^{(k)}$, then it considered as noise because it is not a true unlabeled pixel (i.e., false negative). ADMSE suppress learning from those pixels by assigning small weights during backpropagation. α is a user-defined parameter to control the degree of adaptive weight applied to the loss function. β is a small constant value added to the loss function to ensure stable convergence at the early stage of training (before convergence to correct $\hat{Y}_i^{(k)}$).

3.3 STPR loss

In addition to supervised training using partial labels, additional constraints are leveraged based on prior information of neuronal structures. Because the proposed model is based on multi-task learning using multi-class labels, the model may output strong prediction on more than one class; this is due to the fact that the data and label are noisy and different neuronal structures are overlapped and closely located. The rationale behind the proposed loss is that, for a given pixel location, there should be only one correct class assigned (e.g., a pixel cannot belong to axon and dendrite simultaneously). This constraint can be expressed using conditional probability with random variables $\mathbf{X}$ and $\mathbf{Y}$ as follows:

$$\begin{aligned} P(\mathbf{Y}^{(0)} \cup \mathbf{Y}^{(2)}|\mathbf{X};\theta) &\simeq P(\neg\mathbf{Y}^{(1)} \cap \neg\mathbf{Y}^{(3)}|\mathbf{X};\theta) \\ P(\mathbf{Y}^{(0)} \cup \mathbf{Y}^{(3)}|\mathbf{X};\theta) &\simeq P(\neg\mathbf{Y}^{(1)} \cap \neg\mathbf{Y}^{(2)}|\mathbf{X};\theta) \end{aligned} \tag{8}$$

Intuitively, the above relationship implies that a pixel should not be classified as the cell body or axon if such a pixel has high background or dendrite probability (vice versa for axon as well). Based on this, the STPR loss is proposed as follows:

$$\mathcal{L}_{stpr} = \mathbb{E}_{\mathbf{X}\sim\mathcal{X}} \sum_{k\in S_2} ||\hat{\mathbf{Y}}^{(0)} + \hat{\mathbf{Y}}^{(k)} - \prod_{j\in S(k)} (1 - \hat{\mathbf{Y}}^{(j)})||^2 \tag{9}$$

where $S(k) = \{x|x \in K \setminus \{0, k\}\}$. This loss is applied to noisy labels (i.e., S_2) to serve as a regularizer. In summary, the total loss for our model is defined as the sum of the ADMSE and STPR losses as follows:

$$\mathcal{L}_{total} = \mathcal{L}_{admse} + \mathcal{L}_{stpr} \tag{10}$$

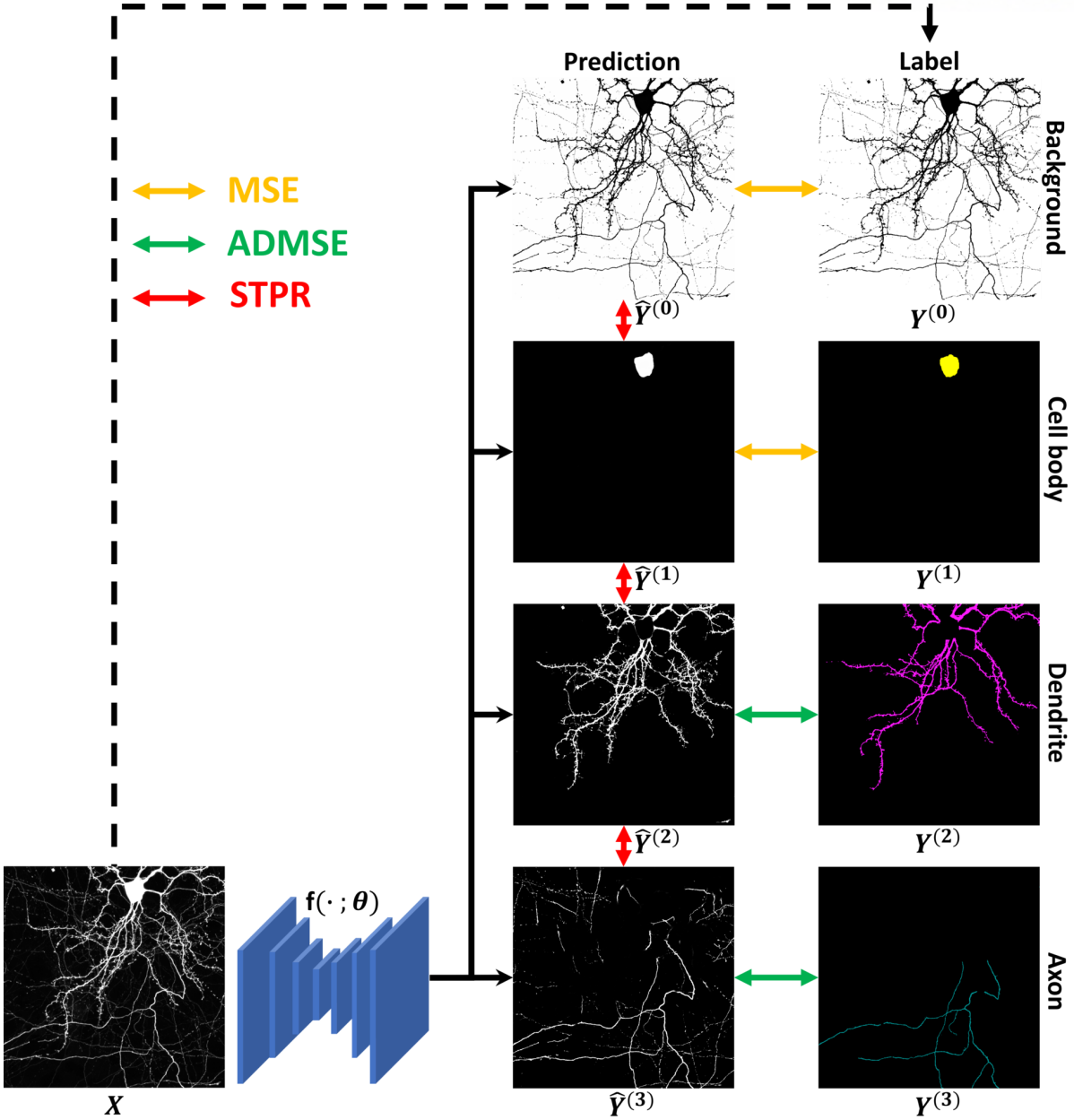


Figure 3: Overview of the proposed method. There are two sets of training data; one is the clean label (background and cell body) and the other is the partial label (dendrite and axon), to which MSE and ADMSE losses are applied, respectively (yellow and green arrows). Predictions from the model is self-regularized using the STPR loss (red arrows).

IV Experiment

4.1 Implementation

Neuron image data used in this thesis research consisted of paired training data with partial annotations from a total of 23 single cells and 3 test sets with full labels. Each image (1024× 1024 in size) was split into small patches (128 × 128 in size). Images are discarded totally empty (background) patches, and applied data augmentation using rotation and flipping. Since the number of pixel labels of dendrite was about five times higher than that of axons, we applied oversampling on axon labels to prevent class imbalanced problem. The proposed network is based on a conventional U-Net with the ResNet-34 backbone encoder. Model is trained by using Adam optimizer with an initial learning rate of 3e-3, with a random batch size of 300, and gradually decreased the learning rate of 3e-4 with the cosine annealing method. U-Net is trained by Resnet-34 as the encoder. Binary image as $\mathbf{Y}^{(0)}$ is generated by empirically setting a threshold of 0.3 for the input image. (see Fig. 3). Loss parameter set $\alpha = 0.1$ to apply the optimal weight map for noisy labels, and $\beta = 0.03$ to converge the learning at initial training, both of which were empirically found via 10-fold cross-validation.

4.2 Experiment Setup

Losses such as MSE, CE and the proposed loss was applied to other networks such as U-net++ [40] and DeepLab V3+ [30]. These networks were set to have the same parameters ($\alpha = 0.1$, $\beta = 0.03$) as the previously set values determined via 10-fold cross-validation. In addition, the images used in this study included not only structural images but also images of mitochondrial features. Each neuron structure has different morphological features and,so much research has been conducted in this images. The proposed method has the advantage of obtaining more mitochondrial features of each structure than the previous algorithm. Therefore, the number of cells is compared through the mitochondria channel and the annotated neuron images. In order to conduct the experiment as Figure 7), the Otsu [41] algorithm was applied to the mitochondrial channel to change a binary images. Then, the Connected component [42] method was applied to the cell count in each structure to obtain the results.

4.3 Evaluation

To demonstrate the efficacy of this method on partial labels, it is compared with other noise-robust losses, such as SCE, APL, and NR-Dice loss. The parameters were $\alpha = 0.1$ and $\beta = 1$ for SCE, $\alpha = 1$ and $\beta = 1$ for APL, and $\gamma = 1$ for NR-Dice, which were chosen empirically to obtain the best performance.

To quantitatively assess the performance of our method, we used three error metrics that are commonly used in semantic segmentation: Intersection over Union (IoU), F1 score, and ClDice [15].

The data are comprised of tubular structures in each neuron; thus, ClDice is employed, which can evaluate the connectivity between structures. To use ClDice to obtain evaluation scores from our structures, the skeletonized ground truth is inserted into the prediction label. More specifically, the values are obtained by skeletonizing the ground truth (V_g) and prediction (V_p) are S_g and S_p, respectively. ClDice value is obtained using precision and recall as shown below.

$$prec(S_p, V_g) = \frac{|S_p \cap V_g|}{|S_p|} \; sens(S_g, V_p) = \frac{|S_g \cap V_p|}{|S_g|}$$

$$ClDice(V_p, V_g) = 2 \times \frac{prec(S_p, V_g) \times sens(S_g, V_p)}{prec(S_p, V_g) + sens(S_g, V_p)}$$

V Result

5.1 Qualitative and Quantitative result

The qualitative comparison results of different noise-robust losses and network with our approach are illustrated in Figure 5,Figure 4. The upper row represents the result of a training image (showing an example of a partial label), while the bottom row indicates the result of a testing image. As shown in this figure, the MSE loss generated incomplete results due to noisy labels. While other losses exhibited some improvement over MSE, none of them was able to generate accurate result close to the full label as in our method (see the red box in Figure 4). Note also that, due to partial labels, other losses failed to learn axons during training (because unlabeled pixels act as a negative label) while our loss successfully reconstructed most of the axons correctly (Figure 4, upper row).

Figure 5 shows that the two proposed losses are applied separately. Compared to the results with the partial labels, if the network is utilized by ADMSE, it can be seen that the unlabeled part is not classified as the background (see the red box in Figure 5). However, since the predicted values are similar, it can be seen that the label is incorrectly annotated. When STPR is applied with MSE, the resulting labels were damaged. This can be inferred occurring because of the unlabeled part contribute during training. As a result, with the two losses, it can be seen that the unlabeled part not only is fully annotated but is also correctly classified under the influence of the regularization through STPR.

Table 3 shows the quantitative comparison of different loss functions with the proposed on, which were obtained by measuring the segmentation accuracy using IoU, precision, recall, F1 score, and ClDice metrics. The last three rows in this table illustrate the ablation study of the proposed loss function. In this result, the axon class is more vulnerable to noisy partial labels than other classes, which resulted in extremely low recall values for the MSE and noise-robust loss functions(between 0.0954 and 0.1676). However, ADMSE and STPR losses are more resilient to label noise and achieved up to 0.5731 for the recall. Results also are observed that dendrites were less affected by noisy labels due to their large and thick structure as compared to thin axons. Therefore, the MSE achieved the best result for IoU (which measures the overlap between two regions) and F1 scores. Although other noise-robust losses demonstrate inferior performance, our proposed loss achieve comparable results to the full label. Moreover, our approach significantly outperformed the other methods in the ClDice metric, which assesses the linearity of the structure, demonstrating that our loss fits better to the neuron segmentation problem than the other noise-robust losses.

5.2 Network result

Figure 6 shows the results showing that is generally effective for other networks. DeepLab V3+ and U-net++ where MSE was applied, so it was shown that the label is not complete for the axon when compared with full labels. (see the large red box in Figure 6). The results demonstrate that, after applying the proposed loss, performance was improved in two networks. In particular, axons are obtained with sufficient labels due to the atrous effect in DeepLab V3+. But from the dendrite, it can be seen that the label in the background pixel is corrupted. In the case of U-net++, which is made with a dense block between the skip connections of the existing U-net, the results is improved compared to MSE.

As shown in Table 2, the evaluation results demonstrate improved performance when applying the MSE and the proposed loss to each network. Since the dendrite area has an error in the background area, the performance is degraded opposite of the axon part in Deep lab V3+. As the proposed loss improves in the recall column compared to MSE, it can be seen that 0.2, 0.03 for each structure is improved, and the F1 score is also improved as a result. In the case of Unet++, it has a structure similar to that of U-net, so improved results can be seen for the dendrite and axon with IOU. Likewise, the F1 score is improved despite the slightly decreased loss of the precision value, and recall is greatly improved by 0.2 and 0.3. In particular, the axon, it was improved by 0.2, and improved performance results were observed. Finally, from evaluating the linear connectivity in each network, in the case of MSE, low results were obtained in the axon region.

5.3 Compare mitochondria

As shown in Figure 9, the ground truth (i.e, full label) shows the results of cell counting for each structure. In the case of mitochondria in the dendrite part, the size of the mitochondria is generally large and they have many long linear characteristics [43]; thus, in a test set with a total of 3 full labels, the number of cells of the proposed method is same as that of MSE. The axons are generally short and have small morphological features [43], so most mitochondria counting was not achieved in MSE, due to many unlabeled parts. On the other hand, the labeling had many instances of successful counting after utilizing the proposed loss. Since the proposed loss significantly improved recall value, it is likely that the axon counts the cells similarly to GT. From this result, I expected that when compared to the traditional method for obtaining mitochondria, the proposed network can be shortened to get mitochondria features and more detail about characteristics in a smaller area.

VI Discussion

In this study, the problem of noise was asymptotic. Previous methods of semantic segmentation were presented that focused on frameworks or preprocessing. This study instead, presented two solutions for noise labeling. First, for stable learning on partial labels, MSE was used as a basic function for unlabeled areas in particular, weights were given based on prediction distance to prevent negative effects. Additionally, to solve the problem of labels being is incorrectly labeled as the background, a structure prior is proposed based on the 'the difference between prediction values suggested in the jigsaw puzzle in self-supervised learning. From the results shown above, it was shown that MAE and MSE used for multi-class image classification are effective in segmentation, and that negative labels can affect learning. Additionally, it was suggested that leveraging to prior information in fluorescence images for learning is an efficient method.

In general, semantic segmentation is approached by multi-class classification in pixels. Unlike the existing semantic segmentation method, this study focused on the false negative area because annotations are required in the unlabeled part. In particular, dendrite is dyed mainly so it can be distinguished in the image. However there is a disadvantage that it is difficult to distinguish because it is not dyed in the axon area. In order to solve this problem, the proposed method reduced false negatives to better detect the axon region.

	IoU			Precision			Recall			F1			ClDice		
	body	Dend	Axon	body	Dend	Axon	body	Dend	Axon	body	Dend	Axon	body	Dend	Axon
MSE	0.8509	**0.7094**	0.0889	0.9138	0.8121	0.6975	0.9251	0.849	0.0954	0.9194	**0.8299**	0.1608	0.8771	0.8105	0.2486
RCE [9]	0.7335	0.6476	0.1719	0.7447	0.7188	0.7661	**0.9799**	0.8677	0.184	0.8461	0.7859	0.2924	0.7321	0.7082	0.4873
APL [10]	0.855	0.6448	0.1304	0.9116	**0.9004**	0.772	0.9325	0.6947	0.1375	0.9218	0.784	0.2293	0.8066	0.7635	0.2851
NR-Dice [11]	0.8374	0.7007	0.1523	0.8707	0.8065	0.6469	0.9566	0.8425	0.1676	0.9115	0.8239	0.2628	0.8215	0.8245	0.3536
$\mathcal{L}_{admse}$(ours)	0.8349	0.6986	0.2871	0.9312	0.7241	0.6084	0.8899	**0.9531**	0.3564	0.9101	0.8223	0.4447	0.8871	0.8101	0.3965
$\mathcal{L}_{mse} + \mathcal{L}_{stpr}$(ours)	0.8626	0.6729	0.2961	0.8848	0.7822	0.7735	0.972	0.8277	0.3241	0.9262	0.8043	0.4568	0.8814	0.7561	0.1683
$\mathcal{L}_{admse} + \mathcal{L}_{stpr}$(ours)	**0.8892**	0.7086	**0.5183**	**0.9676**	0.7435	**0.8439**	0.9165	0.9375	**0.5731**	**0.9413**	0.8292	**0.6826**	**0.9043**	**0.8461**	**0.4948**

Table 1: Comparison of segmentation performance over different loss functions measured using various segmentation quality metrics. The top two results in each case are marked in bold (1st) and underlined (2nd), respectively.

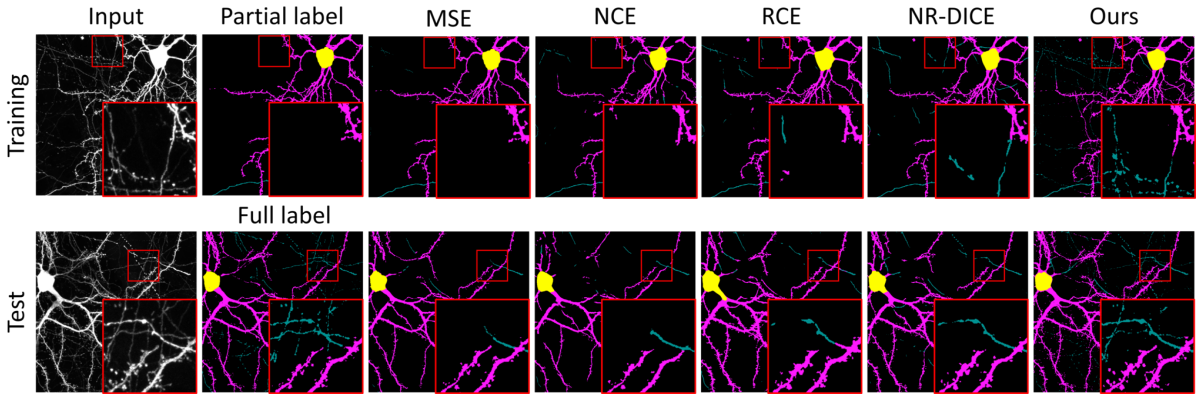


Figure 4: Results of using robust-noise loss functions, such as NCE, RCE, and NR-DICE, as well as the proposed method. Our approach led to detailed annotated results when performing partial label training and demonstrate similar quality training with full label.

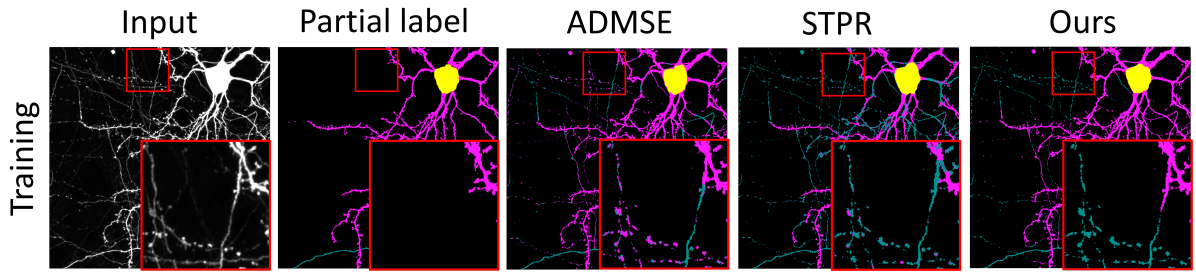


Figure 5: Results of ablation result using proposed method

	IoU			Precision			Recall			F1			ClDice		
	body	Dend	Axon	body	Dend	Axon	body	Dend	Axon	body	Dend	Axon	body	Dend	Axon
Deep lab plus(MSE)	**0.8996**	0.6219	0.2620	**0.9579**	**0.7647**	0.5590	**0.9365**	0.7694	0.3304	**0.9252**	**0.7562**	0.3096	**0.9069**	0.713	0.1612
Deep lab plus(Ours)	0.8438	0.5552	**0.2981**	0.9509	0.5819	**0.611**	0.8879	**0.9236**	**0.3689**	0.9183	0.7141	**0.4593**	0.9033	**0.73**	**0.4011**
Unet ++(MSE)	0.8523	0.6112	0.1953	0.8743	**0.8199**	**0.7048**	**0.9713**	0.7066	0.213	0.9202	0.7585	0.3267	0.8914	0.7924	0.2424
Unet ++(Ours)	**0.8709**	**0.6267**	**0.4235**	**0.9135**	0.6733	0.6654	0.9492	**0.9002**	**0.5404**	**0.9310**	**0.7704**	**0.5950**	**0.9170**	**0.8113**	**0.2704**

Table 2: Comparison of segmentation performance over different network measured using various segmentation quality metrics. The top two results in each case are marked in bold (1st).

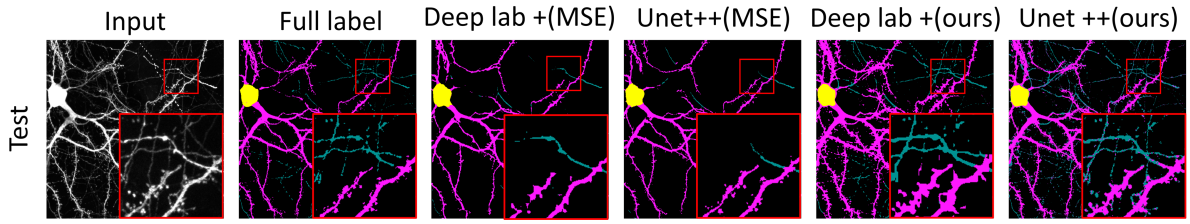


Figure 6: Results of using other segmentation network, such as nested unet, deep lab plus, as well as the proposed method. Our approach led to detailed annotated results when performing MSE loss and demonstrate similar quality with full label.

	Dendrite			Axon		
	GT	**Ours**	**MSE**	**GT**	**Ours**	**MSE**
$Image_1$	127	141	132	46	40	12
$Image_2$	116	117	114	54	47	19
$Image_3$	153	162	154	59	56	14
$Mean$	132	140	133	53	47	15

Table 3: Comparison of a number of cell counting using annotated label based on MSE, proposed loss and GT ,respectively.

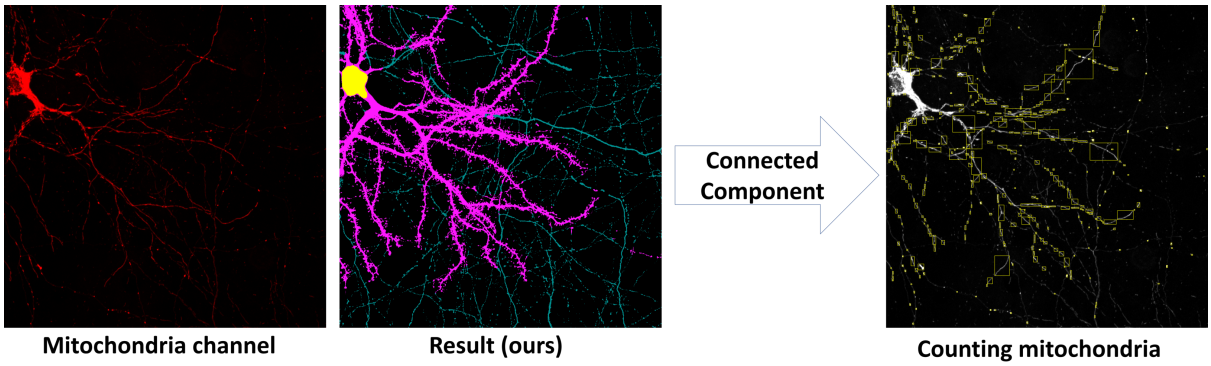


Figure 7: The concept overview of mitochondria counting process. After getting full labeling by using our network, neuron structure label multiple mitochondria channel to counting the mitochondria cell.

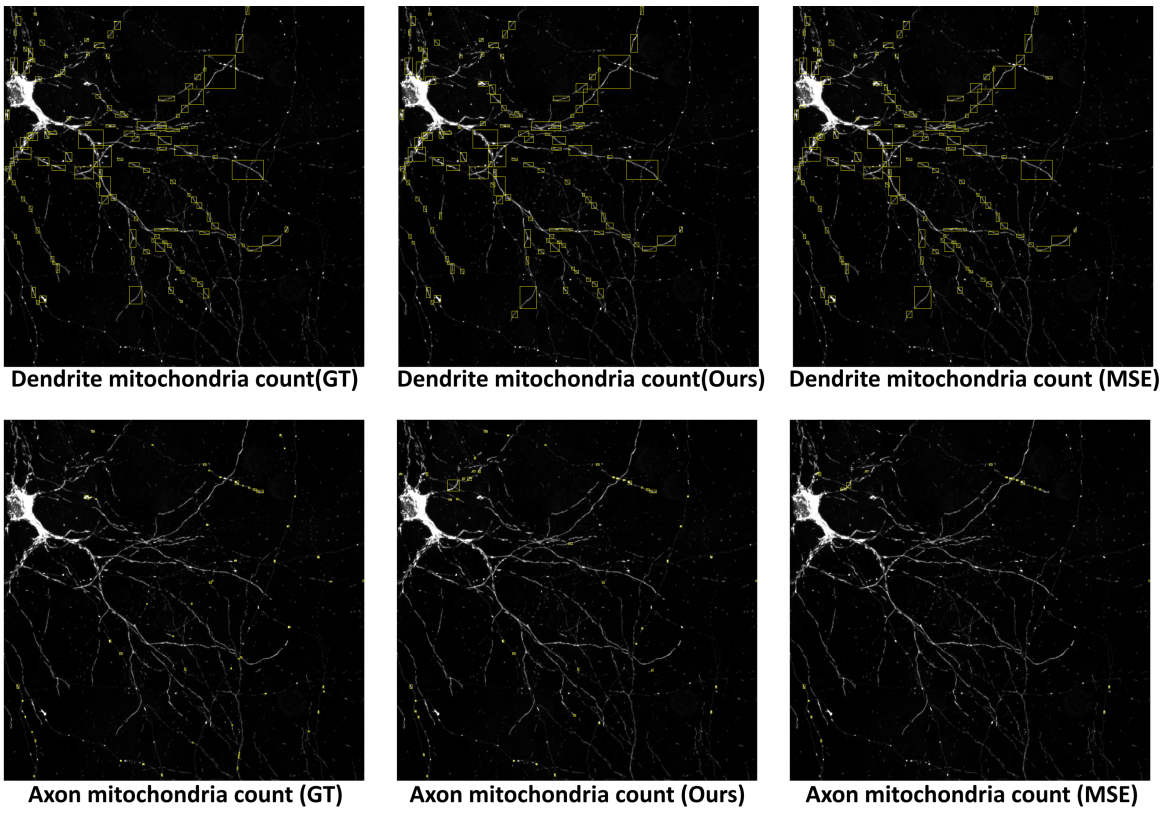


Figure 8: Mitochondria channel can divide two part with axon and dendrite. each structure count number of cell.

VII Conclusion

In this paper, two novel losses are introduced named ADMSE and STPR, for neuron structure segmentation from noisy and incomplete training labels. The result showed that the proposed losses outperformed state-of-the-art noise robust losses in various segmentation quality metrics. Especially, the proposed method achieved best ClDice scores, which demonstrates that the method is effective to the neuron segmentation problem.

The proposed method in this study shows that the noise is partial or incomplete pixel that is labeled as the background. However, corrupted labels caused by actual human error are still unaddressed at this stage of the research. In the future, a more robust loss will be developed to deal with bright images and other types of external noise.

There is disadvantage that fouresacne image must be dyed. This process takes lot of time and becomes more sensitive to intensity. During network learning, it is possible to learn features based on intensity. For these reasons, I will propose method which can learn neuron shape feature in the future work It is expected that network can learn not only the dyed image featuer but also the characteristics of the image itself through a new method.

References

[1] M. London and M. Häusser, "Dendritic computation," *Annu. Rev. Neurosci.*, vol. 28, pp. 503–532, 2005.

[2] J. Grutzendler, K. Helmin, J. Tsai, and W.-B. Gan, "Various dendritic abnormalities are associated with fibrillar amyloid deposits in Alzheimer's disease," *Annals of the New York Academy of Sciences*, vol. 1097, no. 1, pp. 30–39, 2007.

[3] L. Feng, T. Zhao, and J. Kim, "neuTube 1.0: a new design for efficient neuron reconstruction software based on the SWC format," *eneuro*, vol. 2, no. 1, 2015.

[4] H. Zhou, S. Li, A. Li, Q. Huang, F. Xiong, N. Li, J. Han, H. Kang, Y. Chen, Y. Li *et al.*, "GTree: an Open-source Tool for Dense Reconstruction of Brain-wide Neuronal Population," *Neuroinformatics*, pp. 1–13, 2020.

[5] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 3431–3440.

[6] Y. Yuan, X. Chen, and J. Wang, "Object-contextual representations for semantic segmentation," *arXiv preprint arXiv:1909.11065*, 2019.

[7] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional gans," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8798–8807.

[8] A. Ghosh, H. Kumar, and P. Sastry, "Robust loss functions under label noise for deep neural networks," *arXiv preprint arXiv:1712.09482*, 2017.

[9] Y. Wang, X. Ma, Z. Chen, Y. Luo, J. Yi, and J. Bailey, "Symmetric Cross Entropy for Robust Learning With Noisy Labels," *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 322–330, 2019.

[10] X. Ma, H. Huang, Y. Wang, S. Romano, S. Erfani, and J. Bailey, "Normalized Loss Functions for Deep Learning with Noisy Labels," *arXiv preprint arXiv:2006.13554*, 2020.

[11] G. Wang, X. Liu, C. Li, Z. Xu, J. Ruan, H. Zhu, T. Meng, K. Li, N. Huang, and S. Zhang, "A Noise-Robust Framework for Automatic Segmentation of COVID-19 Pneumonia Lesions

From CT Images," *IEEE Transactions on Medical Imaging*, vol. 39, no. 8, pp. 2653–2663, 2020.

[12] Z. Zhang and M. Sabuncu, "Generalized cross entropy loss for training deep neural networks with noisy labels," in *Advances in neural information processing systems*, 2018, pp. 8778–8788.

[13] S. Ruder, "An overview of multi-task learning in deep neural networks," *arXiv preprint arXiv:1706.05098*, 2017.

[14] M. Noroozi and P. Favaro, "Unsupervised learning of visual representations by solving jigsaw puzzles," in *European Conference on Computer Vision*. Springer, 2016, pp. 69–84.

[15] J. C. Paetzold, S. Shit, I. Ezhov, G. Tetteh, A. Ertürk, H. Z. Munich, and B. Menze, "ClDice—a novel connectivity-preserving loss function for vessel segmentation," in *Medical Imaging Meets NeurIPS 2019 Workshop*, 2019.

[16] S. Liu, D. Zhang, Y. Song, H. Peng, and W. Cai, "Automated 3-d neuron tracing with precise branch erasing and confidence controlled back tracking," *IEEE transactions on medical imaging*, vol. 37, no. 11, pp. 2441–2452, 2018.

[17] Z. Zhou, H.-C. Kuo, H. Peng, and F. Long, "Deepneuron: an open deep learning toolbox for neuron tracing," *Brain informatics*, vol. 5, no. 2, pp. 1–9, 2018.

[18] H. Peng, M. Hawrylycz, J. Roskams, S. Hill, N. Spruston, E. Meijering, and G. A. Ascoli, "Bigneuron: large-scale 3d neuron reconstruction from optical microscopy images," *Neuron*, vol. 87, no. 2, pp. 252–256, 2015.

[19] R. Li, T. Zeng, H. Peng, and S. Ji, "Deep learning segmentation of optical microscopy images improves 3-d neuron reconstruction," *IEEE transactions on medical imaging*, vol. 36, no. 7, pp. 1533–1541, 2017.

[20] Q. Li and L. Shen, "3d neuron reconstruction in tangled neuronal image with deep networks," *IEEE Transactions on Medical Imaging*, vol. 39, no. 2, pp. 425–435, 2019.

[21] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *arXiv preprint arXiv:2001.05566*, 2020.

[22] F. Visin, M. Ciccone, A. Romero, K. Kastner, K. Cho, Y. Bengio, M. Matteucci, and A. Courville, "Reseg: A recurrent neural network-based model for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2016, pp. 41–48.

[23] M. Everingham, S. A. Eslami, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes challenge: A retrospective," *International journal of computer vision*, vol. 111, no. 1, pp. 98–136, 2015.

[24] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *European conference on computer vision*. Springer, 2014, pp. 740–755.

[25] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 3213–3223.

[26] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[28] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[29] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1492–1500.

[30] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 801–818.

[31] H. Noh, S. Hong, and B. Han, "Learning deconvolution network for semantic segmentation," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1520–1528.

[32] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[33] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2921–2929.

[34] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618–626.

[35] Z. Huang, X. Wang, J. Wang, W. Liu, and J. Wang, "Weakly-supervised semantic segmentation network with deep seeded region growing," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7014–7023.

[36] D. Lin, J. Dai, J. Jia, K. He, and J. Sun, "Scribblesup: Scribble-supervised convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 3159–3167.

[37] H. Song, M. Kim, D. Park, and J.-G. Lee, "Learning from noisy labels with deep neural networks: A survey," *arXiv preprint arXiv:2007.08199*, 2020.

[38] C. G. Northcutt, L. Jiang, and I. L. Chuang, "Confident learning: Estimating uncertainty in dataset labels," *arXiv preprint arXiv:1911.00068*, 2019.

[39] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," *arXiv preprint arXiv:1710.09412*, 2017.

[40] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer, 2018, pp. 3–11.

[41] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE transactions on systems, man, and cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[42] L. He, Y. Chao, K. Suzuki, and K. Wu, "Fast connected-component labeling," *Pattern recognition*, vol. 42, no. 9, pp. 1977–1987, 2009.

[43] C. E. Park, Y. Cho, I. Cho, H. Jung, B. Kim, J. H. Shin, S. Choi, S.-K. Kwon, Y. K. Hahn, and J.-B. Chang, "Super-resolution three-dimensional imaging of actin filaments in cultured cells and the brain via expansion microscopy," *ACS nano*, 2020.